FORECASTING RESEARCH DIVISION

Technical Report No. 69

Report on the Evaluation of the Satellite Rainfall Algorithms for Nimrod

by

R Brown

September 1993

Met.O.(FR) Technical Report No. 69

Report on the Evaluation of the Satellite Rainfall Algorithms for Nimrod

by

R Brown

September 1993

Met.O.(FR)
(Forecasting Research Division)
Meteorological Office
London Road
Bracknell
Berks, RG12 2SZ

Note

This paper has not been published. Permission to quote from it should be obtained from the Assistant Director of the above Meteorological Office Branch.

Report on the Evaluation of the Satellite Rainfall Algorithms for Nimrod

R Brown

22 September 1993

1 Introduction

A new version of the FRONTIERS satellite precipitation diagnosis step has been developed for use in the automated FRONTIERS component of Nimrod. Evaluation of FRONTIERS forecasts has demonstrated a consistent increase in forecast skill when satellite–derived precipitation is included in the base image. Although it is possible that recent extensions to the radar network may have diminished the impact of the satellite–derived precipitation, it should be noted that it is not uncommon for data from the Shannon radar to be missing. Use of satellite data is then important to give warning of rain approaching from the south west.

The techniques employed by FRONTIERS have been used in the Nimrod version, with the addition of some improvements suggested by the work of Mr Cheng Minghu, a visiting scientist, to optimize their performance. As in FRONTIERS, only the presence or absence of precipitation is diagnosed, not rainfall rate. The possibility of diagnosing limited rainfall rate information is to be examined.

Since the calculation of the satellite precipitation fields is already automated in FRONTIERS (apart from an option to allow the forecasters to define manually the satellite classes associated with precipitation), the areas requiring automation are:–

1. manual registration of the satellite data using any coastline discernible
2. choice of the best satellite precipitation field
3. merging of the satellite and radar data to avoid unrealistic gaps.

No attempt has been made to automate registration with respect coastlines, reliance being placed on the registration performed in Autosat–2. During the trials this proved to be extremely satisfactory.

Optimisation of the satellite diagnosis techniques renders the only choice necessary in (2) to be whether to use a precipitation field derived from visible data only, infrared data only or both together in a two–dimensional (2–D) technique.

Automatic merging of the radar and satellite precipitation fields requires an accurate specification of the range to which the radars can detect precipitation. A theoretical

estimate of radar range, allowing for precipitation depth and beam occultation, the useable data boundary (UDB), is provided by the radar step and is not considered further here. The satellite precipitation field will be used up to the UDB.

The forecaster is allowed to delete or add precipitation to the satellite–derived field. No attempt has been made to automate this process because an overall rainfall analysis is to be developed for Nimrod and this is the logical place for such alterations.

2 The Current FRONTIERS Satellite Step

The FRONTIERS scheme is outlined below because the Nimrod scheme uses the same techniques, although put together in a rather different way.

1. The satellite data are registered manually with reference to any coastline discernible.
2. The satellite data are reprojected from space view to National Grid.
3. The visible data are normalised to a zero zenith angle (Z) using a factor cos(Z). Use of visible data is precluded within about 90 minutes of sunrise and sunset because the simple normalisation scheme breaks down.
4. The Meteosat satellite data and the radar data are compared pixel–by–pixel within the UDB to produce the number of rain and no–rain pixels in each satellite class. These are used to derive the percentage of rain pixels in each satellite class. Tables are produced for visible data alone, infrared data alone (1–D tables) and a 2–D Table. The 1–D tables employ 16 equi–spaced classes to cover the entire range of the raw satellite data. The same classes are used in the 16x16 2–D table.
5. The correlation tables produced every thirty minutes from the individual satellite and radar fields (current tables) are not used directly to assign each satellite class as rain or no–rain. They are used to build "recent" tables which incorporate a degree of time smoothing – recent counts = 0.7 x recent counts + 0.3 x current counts.
6. The satellite classes in the recent table are assigned as rain in decreasing order of percentage of rain pixels until the difference between the diagnosed area and observed area within the UDB is minimised (the "minimum percent" method). This was shown to give more acceptable results subjectively than maximising a skill score such as Hansen and Kuipers or the CSI, Brown and Cheng (1989).
7. Fixed tables (known as universal tables) in which each of the satellite classes is predefined as rain or no–rain are also available.
8. The FRONTIERS operator can define the range of wet visible or infra–red classes manually. The use of this option is hopefully rendered obsolete by optimizing the performance of the correlation technique.

3 The Nimrod Scheme

3.1 Necessary Improvements to the FRONTIERS Scheme

Mr Cheng showed that the 1–D tables require 32 satellite classes, not 16 as used in FRONTIERS. Even with 32 classes, the satellite counts were concentrated in certain classes. The Nimrod scheme uses 32 variable–width classes, designed to avoid bunching of the satellite counts. The 2–D tables still use 16x16 classes but again new variable–width classes have been defined.

If a satellite class does not occur within the radar area it cannot be classified. FRONTIERS assigns unclassified classes as no–rain, no matter how cold and bright they are. The scheme described in Section 3.2 overcomes this problem.

Mr Cheng noted that with a satellite viewing angle around 30^0 the parallax error is significant for high cloud, eg a cloud with top at 10km is placed around 20km (5 pixels) too far north. A correction for parallax error is applied in the Nimrod scheme as developed by Mr Cheng and detailed in FRONTIERS Development Group Tech. Note No. 26.

3.2 Basic Method

The Nimrod scheme uses a mixture of current, recent and universal tables as suggested in FDG Working Paper No. 47. Following the aquisition every half–hour of the satellite data, current and recent correlation tables are constructed as described in Section 2. Universal tables are also available as described in Section 3.3. The satellite data are assigned as rain or no–rain using first the current correlation table. Satellite classes not in the current table are then assigned using the recent table and those not in the recent table are classified using the universal table. When both visible and infrared data are available it is necessary to choose between use of visible alone, infrared alone or the 2–D table. The choice will be made on the basis of which current table exhibits the maximum correlation with the radar data within the UDB.

3.3 Definition of the Universal Tables

By applying the correlation technique to aggregate rain and no–rain count tables, using the minimum percent criterion, Cheng et al. (1993) defined optimum thresholds (1–D tables) and boundaries (2–D tables) to separate rain from no–rain satellite classes for 4 synoptic types. These were used to improve the FRONTIERS universal tables and have also been used to define the Nimrod universal tables. The thresholds and boundaries were fairly similar except for the infrared threshold for cold–air convection. An average boundary for the 2–D table was defined by hand and average thresholds calculated, ignoring the cold–air convection category.

The infrared threshold is -34^0C. The visible threshold chosen initially (103 counts) produced a large underestimate when applied to the first few test cases. This was found to be caused by a shift in the distribution of visible counts from Meteosat in the more recent data compared to the data used by Cheng et al. (1993). The visible threshold was

changed in proportion so that it is now 79 counts. The 2–D boundary was also altered. It was also discovered that the FRONTIERS normalization code, used by both Cheng et al. (1993) and in the Nimrod software, divided the visible counts by 2 after normalization. Therefore in terms of raw Meteosat counts, the visible threshold is 178 counts. According to the expression used to derive albedo in Autosat–2, based on the vicarious calibration of Kriebal and Amann(1990), this is equivalent to an albedo of 55%.

4 Evaluation Procedure

The scheme was evaluated using data from five days. The scheme was run for the period 00z to 23z on each occasion, except where there were problems with corrupt or missing data. Data at hourly intervals were used. Table 1 gives details of the cases used.

The 23 April case was chosen because the precipitation was mainly to the west of the radar area initially, providing a test of the universal tables. However, remnants of the precipitation from a previous front caused a certain number of satellite classes to be classified by correlation. The 24 April was a case of a well–defined band of rain. In contrast, on 3 July about 75% of the radar area was occupied with rain. The 20 July started with a stationary cold front across central areas with scattered light convection to the north. During the afternoon the front decayed as a series of MCS moved NNE from France across SE England.

The test version of the Nimrod software did not select the best satellite precipitation field to use. Visible, infrared and 2–D fields were produced and evaluated in order to examine which statistic should be used to choose the best one. Another version of the software was run which produced separate precipitation fields from the current, recent and universal tables. This was used to check the performance of each, especially the universal tables. The FRONTIERS satellite–derived precipitation was also evaluated both subjectively and objectively. A selection of the FRONTIERS combined radar/satellite fields were examined in order to determine if the forecaster had altered the satellite–derived field. The optimum thresholds at each time were examined for comparison with the thresholds chosen for the universal tables.

The statistics used to evaluate the results, are derived from the following contingency table.

		Precipitation diagnosed	
		yes	no
Precipitation observed	yes	A	B
	no	C	D

Previous experience with the variety of statistics which can be derived from the contingency table suggests that those which agree best with the human perception of an accurate satellite precipitation field are the critical success index (CSI), also known as the threat score, and the tetrachoric correlation coefficient (TCC).

$$CSI = A/(A + B + C) \tag{1}$$

$$TCC = ((A \times D) - (B \times C))/((A + B) \times (A + C) \times (B + D) \times (C + D))^{0.5} \tag{2}$$

CSI values range from 1 for a perfect forecast, to 0 when no precipitation is forecast correctly. The TCC varies from 1 for a perfect forecast to –1 for a perfect anti–forecast (ie all rain pixels forecast as dry and vice versa). The CSI has the advantage that it has been used widely in the nowcasting area. It has the disadvantage that it tends to favour overestimation of rare events and it has a non–zero expectation. The latter means that the average value of CSI obtained by chance is not zero. It's value is particularly dependent on the properties of the observed field. The TCC has the advantage of a zero expectation but has been less widely used than the CSI. Current experience suggests it favours overestimation less than the CSI.

5 Results

5.1 Statistics

Table 2 shows the average TCC, CSI plus expectation (ECSI) and their standard errors. In constructing Table 2 the Nimrod field with the highest TCC was selected at each time when a FRONTIERS field was available. This reproduces how the Nimrod software will work operationally. In each case the Nimrod field achieves a higher average score than that from FRONTIERS, although the difference may not appear statistically significant in terms of the standard error. A simple test of significance would be misleading however because the values of the statistics from Nimrod and FRONTIERS are quite highly correlated. Thus, although the scores for each field vary considerably, the Nimrod and FRONTIERS scores tend to rise and fall together with Nimrod exceeding FRONTIERS. This is illustrated in Figure 1 which shows the Nimrod TCCs plotted against the FRONTIERS TCCs for the individual fields. The 1:1 line is also shown. It can be seen that for the great majority of fields the Nimrod score exceeds the FRONTIERS score. The difference varies considerably and some of the more dramatic differences reflect the uneven performance of the FRONTIERS satellite algorithm software. This is believed to contain at least one bug which it has never been possible to find. The symptom is that the diagnosed fields can suddenly change from being quite good to being large underestimates or large overestimates. An example of a large underestimate is the 20 July 92 field at 22z shown in Figure 6. No such error is apparent in the new software.

Figure 1 shows that occasionally the FRONTIERS field is better than that from the new scheme. One reason may be that it is sometimes advantageous to use the broader classes in FRONTIERS. K. Carpenter, who developed the original scheme, suggested that

for optimum correlation the class boundaries should become variables that are defined for each new field. We have never attempted to produce such a complex scheme.

Table 2 shows that the TCC and CSI do not always agree on the relative skill exhibited by the different cases. For example, the CSI suggests that both FRONTIERS and the Nimrod scheme perform significantly better on 3 July than on 9 June, whilst the TCC suggests no significant difference. The scores can be reconciled by using the difference between the CSI and it's expectation. Very limited subjective assessment has led to the TCC being selected to choose the best field from 2–D, visible or infrared.

Table 3 shows the average TCC values of the current, recent and universal fields for each case. The recent tables score only slightly less than the current tables. Overall the performance of the universal tables is quite good relative to that of the current tables. As pointed out by Cheng and Brown (1991), use of a fixed visible threshold is most successful. The infrared and 2–D universal tables are less accurate because the optimum infrared threshold varies more than the visible one. Table 3 is also consistent with previous studies in that the order of performance on average is 2–D, visible, infrared.

Table 3 does not give a full picture of the advantage of the correlation technique. Use of a fixed threshold leads to larger over or underestimation of the area of precipitation compared to the correlation technique, especially for infrared data. The fact that this has only a modest effect on the TCC probably reflects the limited correlation between precipitation areas and cloud–top properties. Table 3 only applies to evaluation within the training area. The performance of the correlation technique will fall off on average beyond radar range, whilst there should be no trend on average for the universal tables. Thus at some range, which has not been determined, correlation will be no better than the universal tables on average. Use of the correlation technique is still believed worthwhile because:–

(i) the satellite precipitation just beyond radar range will be most important for the forecast

(ii) the work reported in Cheng and Brown (1991) suggests that the correlation technique can work well at long range for the precipitation associated with the same synoptic system as in the radar area. It loses accuracy when a different system is encountered.

(iii) the CPU time is negligible relative to the radar and forecast steps.

5.2 Accuracy of the Universal Table Thresholds

The thresholds produced by the correlation technique have been averaged for comparison with those selected initially for the 1–D universal tables. Only cases when all classes from the coldest or brightest down to the threshold were classified as rain were used because this is how the universal tables are defined. The data from 9 June were excluded because this is a case of scattered convection which requires a different threshold. The mean infrared threshold was -35.3^0C (SE 1.3^0) and the mean visible threshold was 78 counts (SE 1). These are in very good agreement with those selected ie -34^0C and 79 counts.

5.3 Subjective Evaluation

Examples of the satellite diagnosed fields from the Nimrod scheme are shown in Figures 2 – 5 to allow subjective evaluation by the reader. The radar data have been displayed as rain/no–rain only, for ease of comparison with the satellite field. The radar boundary is shown on both radar and satellite fields. The values of TCC are shown in the lower right–hand corner of each satellite field, to aid interpretation of these statistics. The FRONTIERS fields for the cases shown in Figures 2 – 5 are shown in Figure 6.

Some human intervention still appeared necessary on occasions for the best results. The main alteration required was deletion of precipitation diagnosed to be associated with cold cloud well ahead of a warm front or residual cold cloud associated with a decaying system. The FRONTIERS operator deleted precipitation believed spurious at some time during three of the five days studied but was not always consistent in his action. On 23rd April he deleted some precipitation over the North Sea from 03–05z but had not done so at 01z. The precipitation diagnosed to the west of the radar area was rather too extensive and he trimmed it back to agree with the radar for a couple of hours. On 24th April at 02z FRONTIERS produced a massive area of precipitation which the operator trimmed back over the North Sea to agree with the edge of the radar field. At 07z FRONTIERS was still diagnosing some precipitation over the North Sea but this was left in, as it was at 12z (the 12z FRONTIERS field was similar to the 2–D field shown in Figure 2). At 12z precipitation over France was deleted, whilst at 14z this was left in but that over the North Sea was deleted. On 3rd July the operator deleted a thin band of precipitation over the North sea from 00–06z. After 19z, the edge of the diagnosed precipitation over the North Sea was in advance of the edge in the radar field and the satellite field was trimmed back. To be fair to the operators, they are often do not have time to do a thorough job on FRONTIERS and over the North Sea there is little ground truth help decide whether the diagnosed precipitation is real or not.

The Nimrod fields generally contained the features deleted by the operator mentioned above (but not the overdiagnosis on 23rd April). The effect of not deleting these features, apart from early on 23rd April, would not have had a deleterious effect on the forecasts because they were not advected over the UK.

As noted in Cheng et al. (1993), separate correlation tables would have been beneficial when there was a mixture of synoptic systems in the radar area. The 20th July case was a good example of this once the MCS had developed, Fig. 5. The correlation tables were then dominated by the MCS in the SE. Precipitation associated with a decaying cold front NW of the MCS and with weak convection further north was not diagnosed properly because it was associated with a warmer cloud top than the MCS.

Two feature were noticed which might require further action. Despite the routine to exclude visible images at low solar elevation angles, normalization is causing a large increase in brightness for the last few images in the evening. The correlation technique just about copes with this, but it produces massive overdiagnosis from the universal 2–D and visible tables. In the short term, the cut–off data set should be modified. However, an improved normalization scheme would allow more visible data to be used.

The precipitation field occasionally develops holes (a lace curtain effect), generally that from the infrared data, eg Figure 2. This occurs when many adjacent classes have a similar

percentage of rain pixels, so that random fluctuations cause some to be diagnosed as rain and others as no–rain. Some smoothing of the field of percentages may be beneficial. 2

6 Conclusions

The performance of the new satellite rainfall step has been compared with that from FRONTIERS. The scheme is an optimized version of the FRONTIERS scheme and like the latter only diagnoses the presence or absence of precipitation, not rainfall rate.

Objective and subjective assessments were made on five test cases, representing a variety of synoptic situations. Generally, 24 hourly fields were used for each case. Statistical evaluation was by comparison with the radar data.

The statistical evaluation showed that the precipitation field from the new scheme was better that than from FRONTIERS on most occasions. The new scheme performed more reliably than the FRONTIERS scheme. The latter can suddenly produce large over or underestimates, for reasons that it has never been possible to identify. Subjective comparison of the unmodified FRONTIERS satellite precipitation fields with those from the new scheme are in accord with the statistical evaluation.

The FRONTIERS forecaster can modify the satellite precipitation field by adding or deleting precipitation. No attempt has been made to automate this. Such modifications should be made in the overall rainfall analysis being developed for Nimrod. The FRONTIERS forecaster deleted precipitation from a few of the fields from three of the test cases. The precipitation features deleted generally occured in the Nimrod fields, although they were sometimes less extensive. The deletions were generally made to the east of the radar area and would not have been advected over the UK by the forecast step.

7 References

R Brown and M Cheng: Real–time combination of radar and satellite data for very–short–period forecasting. Proc. COST–73 Int. Seminar on Weather Radar Networking, Brussels, 1989, 350–357

M Cheng, R Brown and CG Collier: Delineation of precipitation using infrared and visible data in the vicinity of the United Kingdom. J. Appl. Met., 32, 1993, 884–898

M Cheng: Satellite parallax. FDG Tech. Note No. 26, July 1990

M Cheng and R Brown: Delineation of rain areas using Meteosat IR–VIS data and radar imagery in the region of the United Kingdom. Part 1: Investigation of satellite correlation technique FDG Working Paper No. 46, August 1991

M Cheng and R Brown: Delineation of rain areas using Meteosat IR–VIS data and radar imagery in the region of the United Kingdom. Part 2: A new method. FDG Working Paper No. 47, August 1991

K T Kriebal and V Amann: Absolute calibration of the Meteosat–4 Vis–channel. Proc. 8th Meteosat Scientific Users' Meeting, Norrkoping, Sweden, 1990, 33–38

8 Figure Legend

Figure 1. Values of the tetrachoric correlation coefficient for the individual satellite precipitation fields from the new scheme plotted against those from the FRONTIERS fields. When visible, infrared and 2–D fields were available the largest TCC produced by the new scheme was plotted, since the field producing the largest TCC will be chosen operationally.

Figure 2 – 5. Examples of the satellite–derived precipitation fields on 24 April at 12z, 9 June at 13z, 3 July at 14z and 20 July 1992 at 22z. In each case the 2–D field is shown top left, the infrared field top right, the visible field bottom left and the radar field bottom right. The useable data boundary from FRONTIERS for shallow layer cloud is also shown on each field. TCC values are shown bottom right on each satellite precipitation field derived from comparison with the radar data within the UDB.

Figure 6. Examples of the satellite–derived precipitation fields from FRONTIERS on 24 April at 12z (top left), 9 June at 13z (top right), 3 July at 14z (bottom left) and 20 July 22z (bottom right). TCC values are shown bottom right. The fields were derived from the 2–D correlation table, except on 20 July 92 at 22z when the infrared correlation table was used.

Table 1 Cases Used for Evaluation

DATE	PERIOD	SITUATION
23 April 1992	00 - 23	Initially a cold front, appears to become an instant occlusion later.
24 April 1992	00 - 23	Cold front or occlusion from 23rd crosses country.
9 June 1992	01 - 16	Scattered convection plus some organization into lines and some Cb.
3 July 1992	00 - 23	Widespread rain associated with warm front/cold front combination.
20 July 1992	11 - 23	Cold front, scattered convection, MCS after 16z.

Table 2a Average TCC Values

	23 Apr 92	24 Apr 92	9 Jun 92	3 Jul 92	20 Jul 92
Nimrod	0.35	0.33	0.34	0.36	0.52
FRONTIERS	0.32	0.26	0.30	0.31	0.39
Fields	20	18	9	16	8

Table 2b Average CSI Values and Expectations

	23 Apr 92	24 Apr 92	9 Jun 92	3 Jul 92	20 Jul 92
Nimrod	0.28	0.37	0.25	0.43	0.44
ECSI	0.07	0.19	0.04	0.23	0.10
FRONTIERS	0.26	0.35	0.21	0.38	0.28
ECSI	0.08	0.22	0.03	0.22	0.06

Table 3 Average TCC Values for the Individual Tables

	2-D	Visible	Infrared
Current	0.44	0.36	0.27
Recent	0.41	0.35	0.26
Universal	0.36	0.34	0.22

FIGURE 1

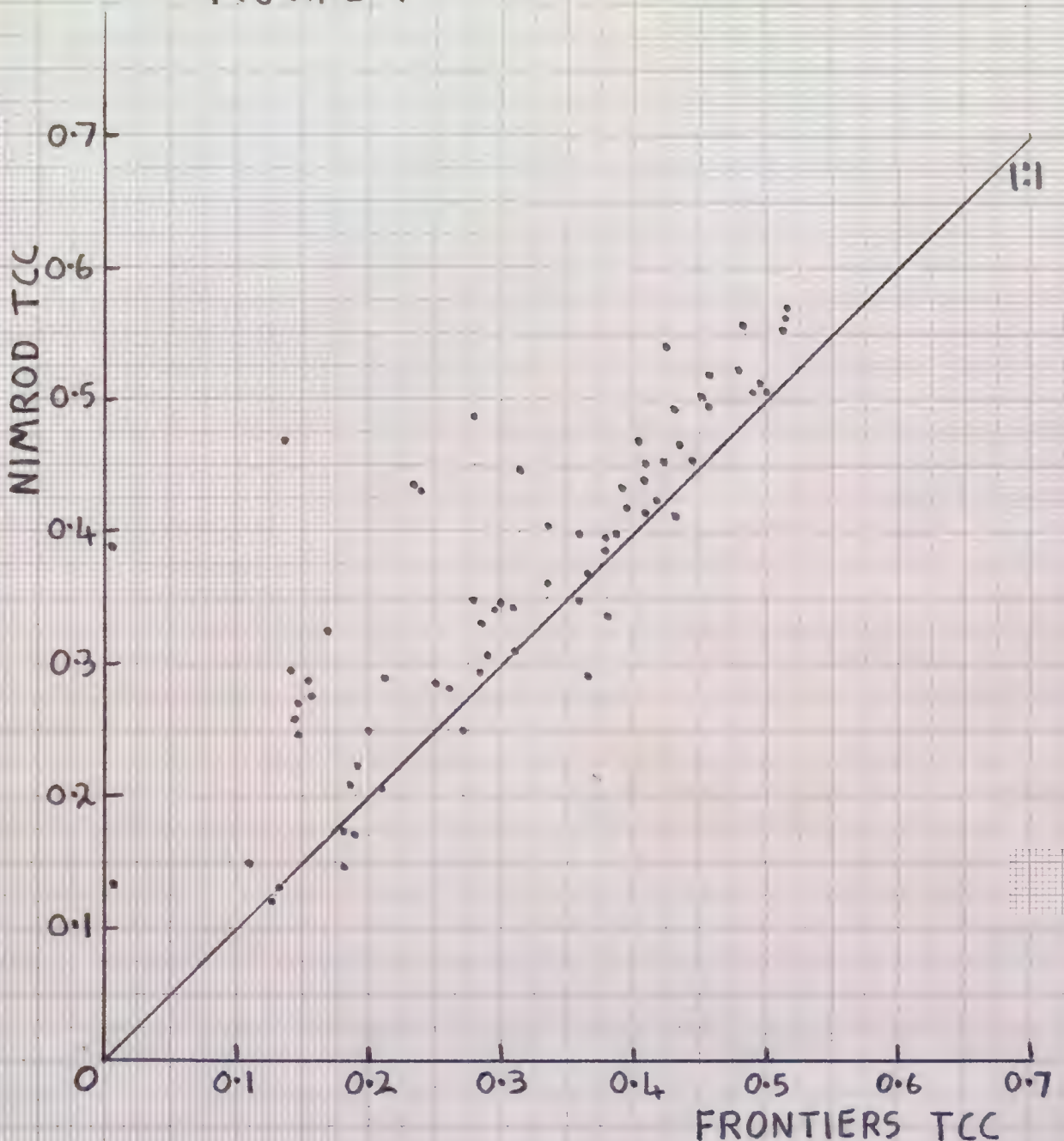

NIMROD TCC
0·7
0·6
0·5
0·4
0·3
0·2
0·1
0
0·1
0·2
0·3
0·4
0·5
0·6
0·7
1:1
FRONTIERS TCC

FIGURE 2

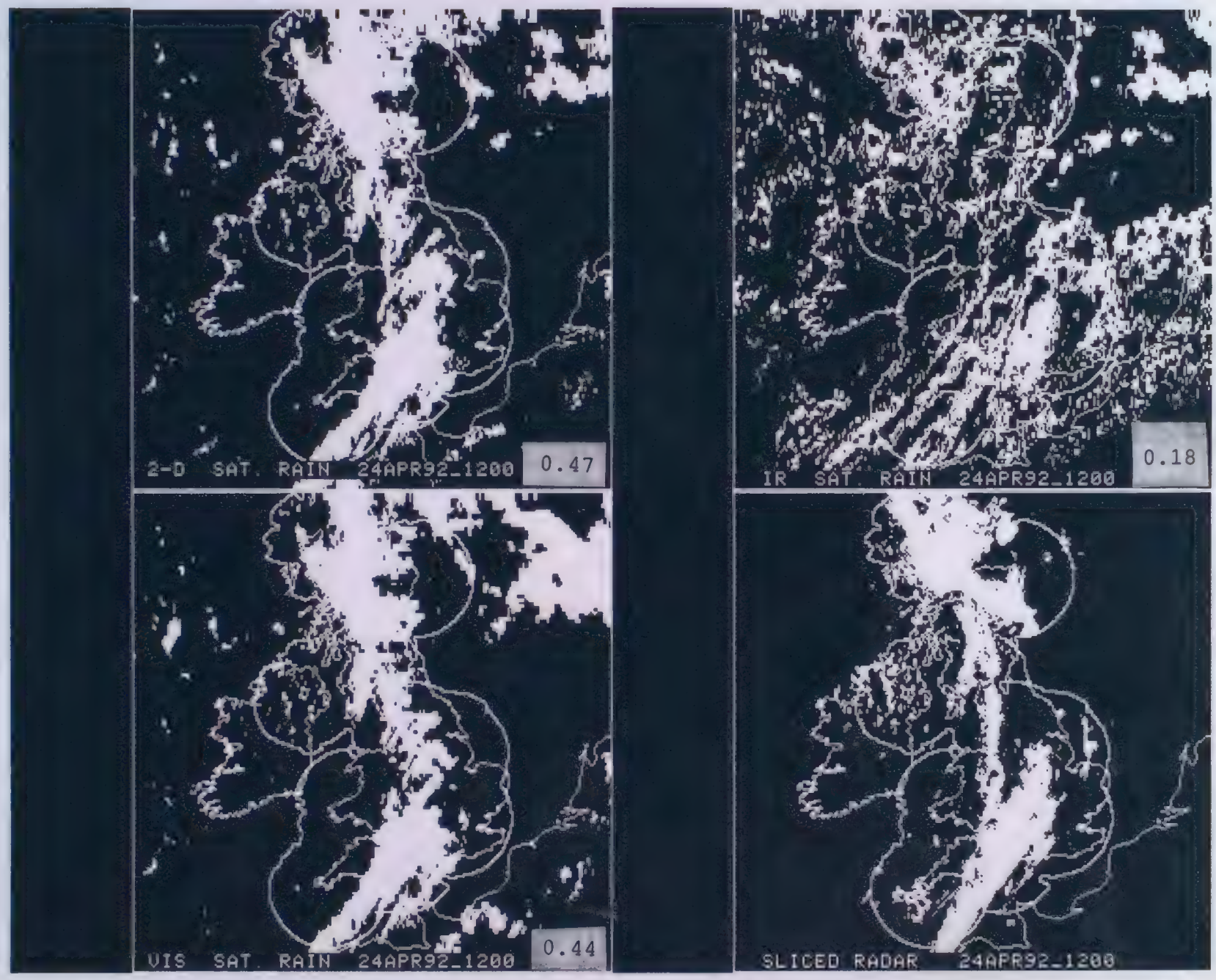

2-D SAT. RAIN 24APR92_1200
0.47
IR SAT. RAIN 24APR92_1200
0.18
VIS SAT. RAIN 24APR92_1200
0.44
SLICED RADAR 24APR92_1200

FIGURE 3

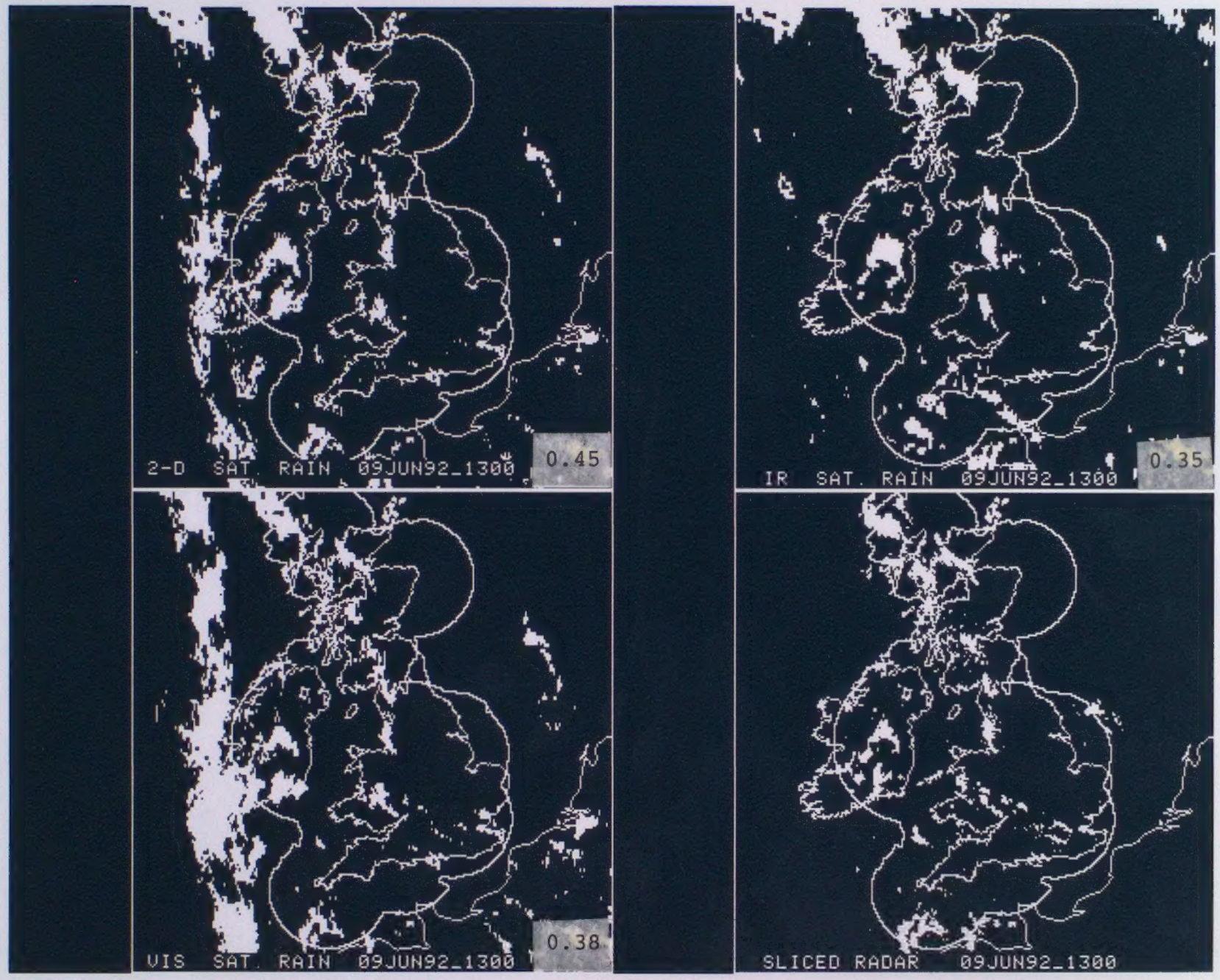

2-D SAT. RAIN 09JUN92_1300
0.45
IR SAT. RAIN 09JUN92_1300
0.35
VIS SAT. RAIN 09JUN92_1300
0.38
SLICED RADAR 09JUN92_1300

FIGURE 4

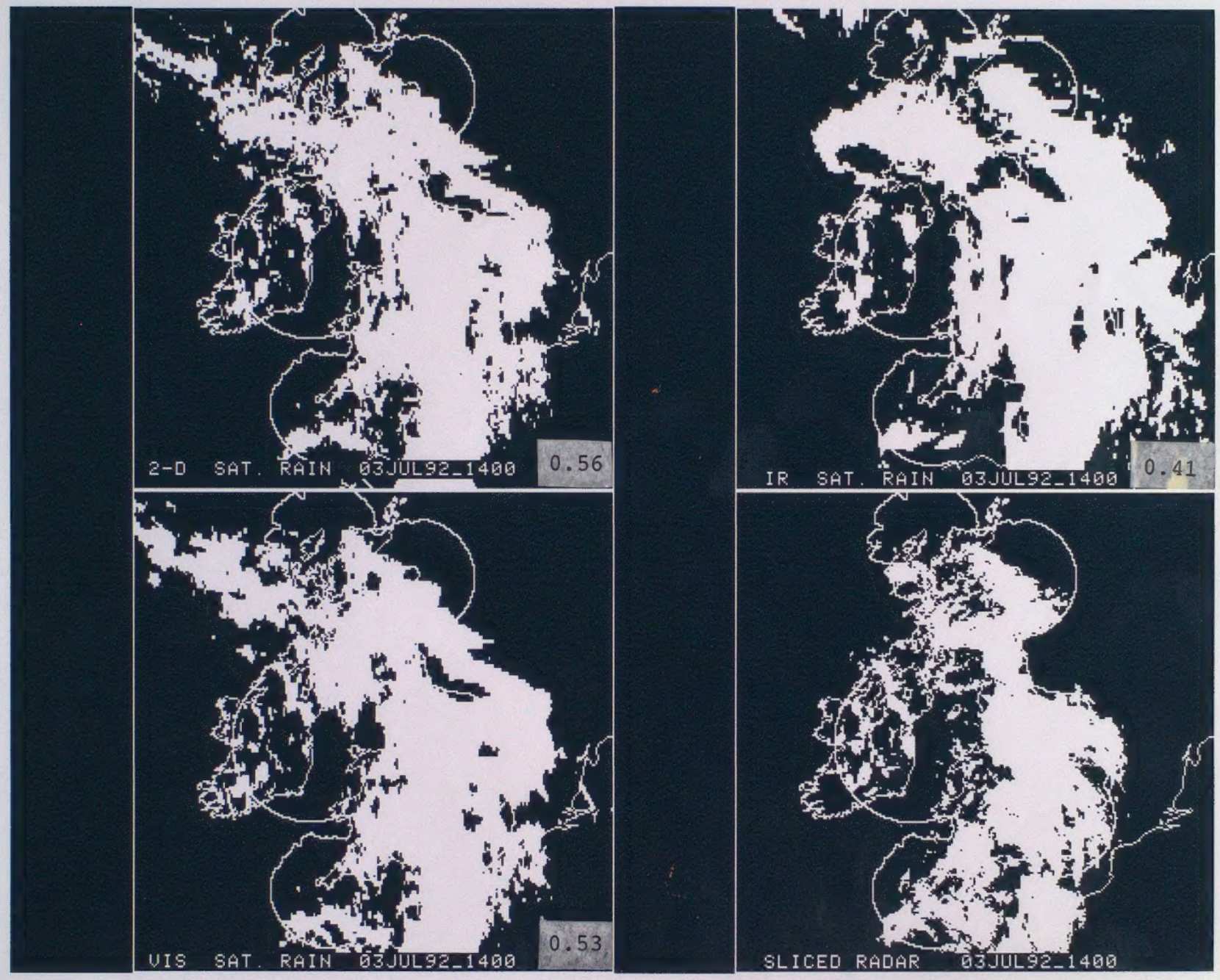

2-D SAT. RAIN 03JUL92_1400
0.56
IR SAT. RAIN 03JUL92_1400
0.41
VIS SAT. RAIN 03JUL92_1400
0.53
SLICED RADAR 03JUL92_1400

FIGURE 5

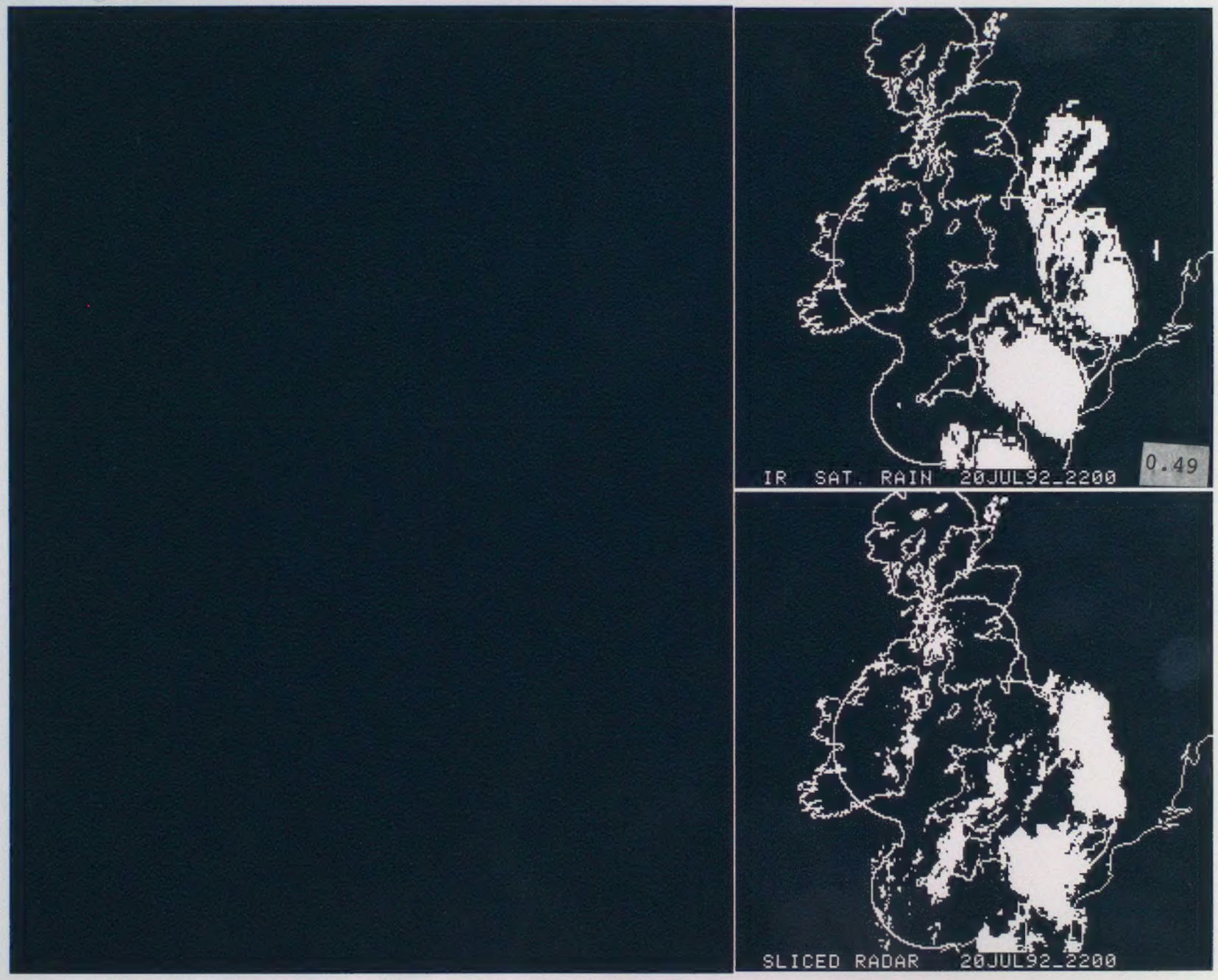
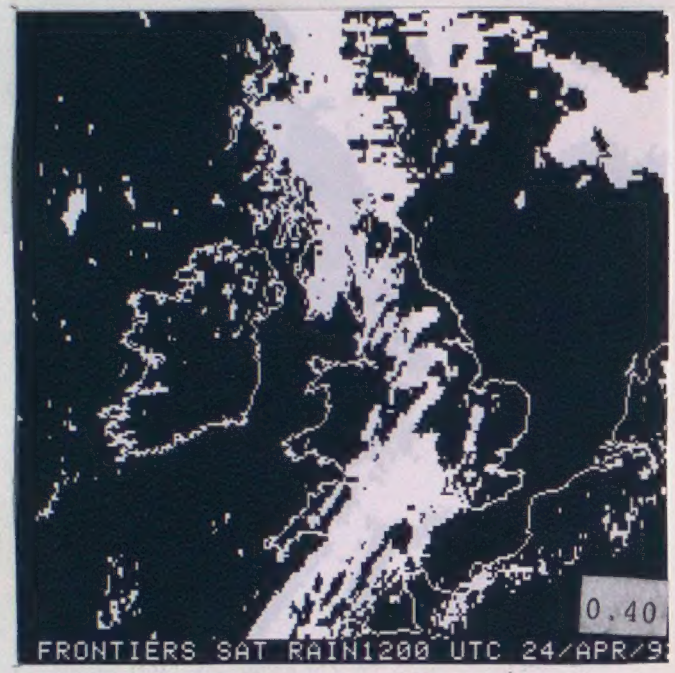


FIGURE 6


0.40
FRONTIERS SAT RAIN1200 UTC 24/APR/9

0.38
FRONTIERS SAT RAIN1300 UTC 09/JUN/9

0.48
FRONTIERS SAT RAIN1400 UTC 03/JUL/9

0.28
FRONTIERS SAT RAIN2200 UTC 20/JUL/9